DOCUMENT RESUME

ED 380 507 TM 022 874

AUTHOR Hester, Yvette
TITLE Least Squares Procedures.
PUB DATE Jan 95
NOTE 24p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Dallas, TX, January 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Goodness of Fit; *Least Squares Statistics; *Matrices; *Regression (Statistics)
IDENTIFIERS Nonlinear Models; *Parametric Analysis

ABSTRACT

Least squares methods are sophisticated mathematical curve fitting procedures used in all classical parametric methods. The linear least squares approximation is most often associated with finding the "line of best fit" or the regression line. Since all statistical analyses are correlational and all classical parametric methods are least square procedures, it becomes imperative to understand just what the least squares procedure is and how it works. This paper illustrates the least squares procedure, starting with one independent variable and one dependent variable and generalizes to "n" independent variables with vector and matrix notation. Graphical representations and small heuristic examples are given. A brief generalization to nonlinear squares is presented. (Three tables and three figures illustrate the analysis. An appendix gives software commands for analysis. Contains 3 references.) (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

YVETTE HESTER

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Least Squares Procedures

Yvette Hester

Texas A&M University

ED 380 507

Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, TX, January 26, 1995

TM022874

Abstract

Least squares methods are sophisticated mathematical curve fitting procedures used in all classical parametric methods. The linear least squares approximation is most often associated with finding the "line of best fit" or the regression line. Since all statistical analyses are correlational and all classical parametric methods are least squares procedures, it becomes imperative to understand just exactly what the least squares procedure is and how it works.

This paper illustrates the least squares procedure, starting with the simplest case of linear least squares with one independent variable and one dependent variable and generalizes to n independent variables with vector and matrix notation. Graphical representations and small heuristic examples are given. A brief generalization to nonlinear least squares is presented.

The least squares method is a mathematical curve fitting procedure. The linear least squares approximation is most often associated with finding the "line of best fit" or the regression line. Since all statistical analyses are correlational and all classical parametric methods are least squares procedures (Thompson, 1994), it becomes imperative to understand just exactly what the least squares procedure is and how it works.

The purpose of this paper is to illustrate the least squares procedure, starting with the simplest case of linear least squares with one independent variable and one dependent variable, and generalizing to cases which better model reality. Algebra, matrix algebra and calculus will be employed to some extent, with explanation at each step for following the logic. Examples with small data sets are given to help make the procedure more concrete.

Linear Least Squares

In the simplest case, the linear relationship between one independent variable and one dependent variable is considered. Thinking conceptually about correlation, the question might be posed, "How well does a line catch all the points in a scattergram of data?". Thinking conceptually about the line catching those points, the question is, "How can the equation of the line that does the best at catching those points be found?". This line is the least squares line in which the sum of the squares of the vertical distances from the data points to the line are made as small as possible or minimized. The line must be the best fit for all the points simultaneously. Figure 1 illustrates a line closest to four data points and the distances whose sum of squares is to be minimized.

Insert Figure 1 here

Note that if all the points were to lie exactly on a straight line, any two of those points could be used to determine the equation of the line using the point-slope form $y - y_1 = b(x - x_1)$, where (x_1, y_1) is any one of the points on the line and b is the slope of the line given by $b = \frac{y_2 - y_1}{x_2 - x_1}$, with (x_2, y_2) as another point on the line. Since all points almost never lie exactly on a straight line, the least squares procedure is invoked to determine the equation of the line of best fit.

To generalize from the distances illustrated in figure 1, suppose that there are n data points and that a somewhat linear relationship is expected. The least squares line can be found by minimizing the sum

(1) $$(d_1)^2 + (d_2)^2 + \ldots + (d_n)^2 = \sum_{i=1}^{n} d_i^2.$$

The equation of the regression line is given by $y' = bx + a$, where b is the slope of the line and a is the y-intercept. The y-values lying on the regression line corresponding to a particular x_i are denoted y_i' since they are the predicted values and not the observed y-values for that x. In other words, the y'-values do not correspond to the points that actually appear in the scattergram, unless the line catches them exactly, but are the y-coordinates of the points on the line. Each vertical distance is the difference in the y' and the y values. Using the substitution $y' = bx + a$, each distance can be written as,

$$d_i = y_i' - y_i = bx_i + a - y_i, \text{ for all } \leq i \leq n.$$

The difference can be written as $y_i' - y_i$ or as $y_i - y_i'$ without changing the following results. The substitution changes equation (1), the sum to be minimized, to

(1)' $$(bx_1 + a - y_1)^2 + (bx_2 + a - y_2)^2 + \ldots + (bx_n + a - y_n)^2 = \sum_{i=1}^{n} (bx_i + a - y_i)^2.$$

The ordered pairs $(x_1,y_1),(x_2,y_2),...,(x_n,y_n)$ are all known, since they are the actual data points. The unknown values b and a in the equation are left to be found.

A calculus technique will be used to minimize equation (1)'. Since there are two unknown variables, b and a, in the equation, partial derivatives will be taken and set equal to zero to solve. This technique from calculus will allow us to both minimize the one equation and solve for the missing values using linear algebra. If the following assumptions are met, then the ordinary least squares approximation is the best (most efficient) linear unbiased estimator (BLUE). The four assumptions are (1) fixed x-values, (2) homoscedasticity, (3) error terms are uncorrelated (no autocorrelation), and (4) error terms have zero mean (Hamilton, 1992).

For ease of notation, let $(1)' = f$. The partial derivatives of f are

$$\frac{\partial f}{\partial b} = 2(bx_1 + a - y_1)(x_1) + 2(bx_2 + a - y_2)(x_2) + ... + 2(bx_n + a - y_n)(x_n)$$

$$= 2\sum_{i=1}^{n}(x_i)(bx_i + a - y_n)$$

$$\frac{\partial f}{\partial a} = 2(bx_1 + a - y_1) + 2(bx_2 + a - y_2) + ... + 2(bx_n + a - y_n)$$

$$= 2\sum_{i=1}^{n}(bx_i + a - y_i).$$

Combining like terms and setting the partials equal to zero:

$$\frac{\partial f}{\partial b} = 2\left(b\sum_{i=1}^{n}x_i^2 + a\sum_{i=1}^{n}x_i - \sum_{i=1}^{n}x_iy_i\right) = 0$$

$$\frac{\partial f}{\partial a} = 2\left(b\sum_{i=1}^{n}x_i + na - \sum_{i=1}^{n}y_i\right) = 0 .$$

Writing the partials in simplest form yields a system of two equations and two unknowns.

$$\text{(i)} \qquad b\left(\sum_{i=1}^{n} x_i^2\right) + a\left(\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} x_i y_i$$

$$\text{(ii)} \qquad b\left(\sum_{i=1}^{n} x_i\right) + na = \sum_{i=1}^{n} y_i$$

Using the method of elimination by addition for solving systems and the multiplicative constants $-n \cdot (i)$ and $\sum_{i=1}^{n} x_i \cdot (ii)$, one equation is obtained.

$$(i)' \qquad -nb\left(\sum_{i=1}^{n} x_i^2\right) - na\left(\sum_{i=1}^{n} x_i\right) = -n\left(\sum_{i=1}^{n} x_i y_i\right)$$

$$(ii)' \quad + \qquad b\left(\sum_{i=1}^{n} x_i\right)^2 + na\left(\sum_{i=1}^{n} x_i\right) = \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$$

$$b\left(\sum_{i=1}^{n} x_i\right)^2 - nb\left(\sum_{i=1}^{n} x_i^2\right) = \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) - n\left(\sum_{i=1}^{n} x_i y_i\right)$$

To solve the resulting equation for b, factor b out of the first two terms and divide.

$$b\left[\left(\sum_{i=1}^{n} x_i\right)^2 - n\left(\sum_{i=1}^{n} x_i^2\right)\right] = \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) - n\left(\sum_{i=1}^{n} x_i y_i\right)$$

$$\text{(2)} \qquad b = \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) - n\left(\sum_{i=1}^{n} x_i y_i\right)}{\left(\sum_{i=1}^{n} x_i\right)^2 - n\left(\sum_{i=1}^{n} x_i^2\right)}$$

Solving (ii) for a yields:

$$(3) \qquad a = \frac{\sum_{i=1}^{n} y_i - b\left(\sum_{i=1}^{n} x_i\right)}{n}, \quad \text{where b is as above.}$$

The linear least squares equation or the regression equation for n data points is $y' = bx + a$, where b and a satisfy equations (2) and (3) derived above, and yields the line of best fit for a particular data set. Equations (2) and (3) above are referred to as the normal equations. As the normal equations suggest, find b first, then use (3) to find a.

Example 1

Suppose a teacher is interested in the relationship between the number of classes a student has missed over a semester and the grade the student received in the class. The data are found in table 1.

Insert Table 1 here

The number of absences is represented by the independent variable x and the grade for the course is represented by the dependent variable y. Figure 2 illustrates the scattergram of the data.

Insert Figure 2 here

To find the least squares equation for this example, equations (2) and (3) are used to calculate b and a. The calculations are presented in table 2.

Insert Table 2 here

Substituting from table 2 into equation (2),

$$b = \frac{702 - 480}{676 - 1200} = \frac{222}{-524} \cong -.4.$$

Since the slope of the line is negative, the graph of the line will fall from left to right. The scattergram suggests that a falling line is appropriate. Substituting $b = -.4$ and values from table 2 into equation (3),

$$a = \frac{27 + 10.4}{10} = 3.74.$$

The y-intercept of the line is 3.74 and the equation of the least squares line or regression line is $y' = 3.74 - .4x$. The least squares line is illustrated on the scattergram in figure 3.

Insert Figure 3 here

<u>The Matrix/Vector Approach</u>

In order to move to a more general case, the notation of vectors and matrices is adopted. The simple case already considered will be adapted to the new notation for clear understanding. Taking the general equation for the vertical distances $d_i = y_i' - y_i = bx_i + a - y_i$, each distance can be written specifically as

$$\begin{cases} a + bx_1 - y_1 \\ a + bx_2 - y_2 \\ \vdots \\ a + bx_n - y_n \end{cases}$$

This set of equations decomposes into the matrix $A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$ and the vectors $u = \begin{bmatrix} a \\ b \end{bmatrix}$ and $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$. Then the linear least squares procedure finds u, and therefore b and a, so that $|Au - y|$ is minimized.

Matrices have dimension (row by column). The matrix A has dimension n by 2, written n x 2. The column matrix or vector u has dimension 2 x 1 and the column matrix or vector y has dimension n x 1. The distance $|Au - y|$ suggests the matrix multiplication of A and u. This product matrix can be found because the matrices satisfy the linear algebra rule that the number of columns of the first matrix must equal the number of rows of the second matrix. The resulting product will also be a matrix and will have dimension equal to the number of rows of the first matrix by the number of columns of the second matrix. In this case, that resulting matrix would have dimension n x 1. In order to subtract two matrices, which is the operation that must be performed next, the two matrices must have exactly the same dimension. Since the column matrix y is $n\,x$, it is important to the procedure for the resulting product matrix Au to also have dimension n x 1, which it does, as shown above.

A more general linear least squares case would involve two independent variables and a dependent variable. The data points or ordered triples for n observations would look like $(x_{11}, x_{21}, y_1), (x_{12}, x_{22}, y_2), \ldots, (x_{1n}, x_{2n}, y_n)$. The researcher might wonder whether the y values are linearly related to the x values. Each distance would be written

$$\begin{cases} b_0 + b_1 x_{11} + b_2 x_{21} - y_1 \\ b_0 + b_1 x_{12} + b_2 x_{22} - y_2 \\ \vdots \\ b_0 + b_1 x_{1n} + b_2 x_{2n} - y_n \end{cases}$$

The matrix and vectors for this case generalize to

$$A = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{bmatrix} \qquad B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

(n x 3) (3 x 1) (n x 1)

The linear least squares procedure will then determine the vector B (and hence b_0, b_1, b_2) by again minimizing the difference $|AB - Y|$. Note that the multiplication can be performed since A is (n x 3) and u is (3 x 1). The resulting product matrix has dimension (n x 1).

The extension to the most general linear least squares case of one dependent and k independent variables follows the same notation and will offer the generalized normal equation(s). The linear least squares approximation now has the form

$$y_i' = b_0 + b_1 x_{k1} + b_2 x_{k2} + \ldots + b_k x_{kn}.$$

The distances are now

$$\begin{cases} b_0 + b_1 x_{11} + b_2 x_{21} + \ldots + b_k x_{k1} - y_1 \\ b_0 + b_1 x_{12} + b_2 x_{22} + \ldots + b_k x_{k2} - y_2 \\ \vdots \\ b_0 + b_1 x_{1n} + b_2 x_{2n} + \ldots + b_k x_{kn} - y_n \end{cases},$$

with the following matrix decomposition,

$$A = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}_{(n \times (k+1))} \qquad B = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}_{((k+1) \times 1)} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{(n \times 1)}.$$

Checking dimensions for multiplication of the matrices yields AB as an (n x 1) product matrix. The linear least squares procedure will minimize the difference $|AB - Y|$. If the columns of A are linearly independent, then the difference $(Y' - Y)$ is uniquely determined, $A'A$ is positive definite (and therefore, nonsingular), (Seber, 1977), $(A'A)^{-1}$ exists and

(*) $$B = (A'A)^{-1} A'Y,$$

where A' denotes the transpose of A. The general normal equation(s) above minimizes the difference $|AB - Y|$. An indepth discussion of terms used in the above reference and the derivation of (*) are beyond the scope of this paper.

Example 2

Suppose a teacher is interested in whether math SAT scores, scores on a high school math achievement test, and the grade received in a first college math class are linearly related to college grade point averages for mathematically gifted students. The data for 10 hypothetical students are found in table 3.

Insert Table 3 here

The independent variable X_1 contains the ten math SAT scores, the independent variable X_2 contains the ten high school math achievement test scores, the independent variable X_3 contains the ten first math course grades and the student grade point averages are represented by the dependent variable Y. The matrix representation for this example is

$$A=\begin{bmatrix}1 & 670 & 10 & 4\\ 1 & 630 & 9 & 4\\ 1 & 610 & 9 & 4\\ 1 & 570 & 9 & 3\\ 1 & 700 & 8 & 3\\ 1 & 640 & 6 & 2\\ 1 & 630 & 7 & 2\\ 1 & 610 & 8 & 3\\ 1 & 570 & 10 & 4\\ 1 & 550 & 8 & 2\end{bmatrix} \quad Y=\begin{bmatrix}3.82\\ 3.74\\ 1.84\\ 3.34\\ 3.26\\ 2.35\\ 3.03\\ 3.05\\ 3.76\\ 2.72\end{bmatrix} \quad B=\begin{bmatrix}b_0\\ b_1\\ b_2\\ b_3\end{bmatrix}$$

where B is the matrix of constants to be found by the linear least squares procedure. The difference to be minimized is

$$|AB-Y|=\left|A_{(10x4)}B_{(4x1)}-Y_{(10x1)}\right| =\left|(\mathrm{AB})_{(10x1)}-Y_{(10x1)}\right|.$$

The column matrix or vector B is found by using equation (*). Since the inverse of $A'A$ must be found, the computer algebra system MAPLE will be implemented for computational ease. The inverse of a (2 x 2) or (3 x 3) matrix can easily be computed by hand, but since $A'A$ has dimension (4 x 4) the computations are best left to the computer. As equation (*) is implemented, the step by step matrix products will be given.

Equation (*) requires first that the transpose of A be taken. The transpose of any matrix is found by switching the rows and the columns. Let the columns of A be the rows of A' and the rows of A be the columns of A'. Then,

$$A' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 670 & 630 & 610 & 570 & 700 & 640 & 630 & 610 & 570 & 550 \\ 10 & 9 & 6 & 9 & 8 & 6 & 7 & 8 & 10 & 8 \\ 4 & 4 & 1 & 3 & 3 & 2 & 2 & 3 & 4 & 2 \end{bmatrix}.$$

Next, the product $A'A$ must be found. Note that since A' has dimension (4 x 10) and A has dimension (10 x 4), the product can be formed and the resulting matrix will have dimension (4 x 4). To perform matrix multiplication, "pour" the rows of the first matrix down the columns of the second, multiply like entries and add for a total entry.

$$A'A = \begin{bmatrix} 10 & 6{,}180 & 81 & 28 \\ 6{,}180 & 3{,}838{,}800 & 49{,}990 & 17{,}370 \\ 81 & 49{,}990 & 675 & 239 \\ 28 & 17{,}370 & 239 & 88 \end{bmatrix}$$

The inverse of the above (4 x 4) matrix must be found. Recall that the inverse must exist because the columns of the product matrix are linearly independent and the matrix itself is positive definite. MAPLE finds

$$(A'A)^{-1} = \begin{bmatrix} \frac{2{,}376{,}803}{40{,}143} & \frac{-26{,}141}{401{,}430} & \frac{-56{,}867}{13{,}381} & \frac{223{,}069}{40{,}143} \\ \frac{-26{,}141}{401{,}430} & \frac{163}{2{,}007{,}150} & \frac{243}{66{,}905} & \frac{-2{,}077}{401{,}430} \\ \frac{-56{,}867}{13{,}381} & \frac{243}{66{,}905} & \frac{6{,}114}{13{,}381} & \frac{-8{,}104}{13{,}381} \\ \frac{223{,}069}{40{,}143} & \frac{-2{,}077}{401{,}430} & \frac{-8{,}104}{13{,}381} & \frac{36{,}506}{40{,}143} \end{bmatrix}.$$

To find the inverse of a matrix by hand, augment the matrix with the identity matrix of the same dimension and perform Gauss-Jordan elimination. This process can get messy as demonstrated by the fractional entries above.

The first half of equation (*) has now been found. To proceed, the product of the inverse matrix and A' is found. The product will combine a (4 x 4) matrix and a (4 x 10) matrix, to form a (4 x 10) matrix. Even simple

operations such as matrix multiplication can become complicated. The computer algebra system MAPLE computed the following product matrix:

$$(A'A)^{-1}A' = \begin{bmatrix} \frac{-188,378}{40,143} & \frac{28,929}{13,381} & \frac{-18,335}{40,143} & \frac{20,564}{40,143} & \frac{-49,556}{13,381} & \frac{126,311}{40,143} & \frac{-18,149}{40,143} & \frac{28,867}{13,381} & \frac{24,344}{13,381} & \frac{20,378}{40,143} \\ \frac{1,973}{401,430} & \frac{-263}{133,810} & \frac{208}{200,715} & \frac{-334}{200,715} & \frac{352}{66,905} & \frac{-683}{401,430} & \frac{449}{401,430} & \frac{-137}{66,905} & \frac{-429}{133,810} & \frac{-701}{401,430} \\ \frac{4,419}{13,381} & \frac{-3,639}{13,381} & \frac{1,549}{13,381} & \frac{1,753}{13,381} & \frac{-5,287}{13,381} & \frac{341}{13,381} & \frac{-2,621}{13,381} & \frac{-441}{13,381} & \frac{2,567}{13,381} & \frac{28,929}{13,381} \\ \frac{-13,186}{40,143} & \frac{6,478}{13,381} & \frac{-12,994}{40,143} & \frac{-4,610}{40,143} & \frac{-2,433}{13,381} & \frac{17,281}{40,143} & \frac{-4,954}{40,143} & \frac{3,798}{13,381} & \frac{2,528}{13,381} & \frac{-12,650}{40,143} \end{bmatrix}$$

The last product to be formed is that of the above matrix with the matrix Y. The result will be the matrix product $(A'A)^{-1}A'Y$ which contains the unknown constants in the column matrix B and the solutions to the linear least squares normal equation(s).

$$B = \begin{bmatrix} -.21747 \\ .001265 \\ .194196 \\ .340553 \end{bmatrix}$$

Thus, the linear least squares equation for example 2 is

$$y' = -.21747 + .001265x_1 + .194196x_2 + .340553x_3.$$

Nonlinear least squares

A brief discussion without examples of quadratic least squares minimum differences and general least squares minimum differences will be related to the linear least squares discussion above.

If data represented by the ordered pairs $(x_1, y_1), (x_2 y_2), \ldots (x_n, y_n)$ has been collected and the researcher expects a quadratic relationship between the x and y values, the distances are represented as

$$\begin{cases} b_0 + b_1 x_1 + b_2 x_1^2 - y_1 \\ b_0 + b_1 x_2 + b_2 x_2^2 - y_2 \\ \vdots \\ b_0 + b_1 x_n + b_2 x_n^2 - y_n \end{cases}$$

Consequently, the matrix decomposition looks like:

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \qquad B = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

The difference to be minimized is the same $|AB - Y|$, where B is found through the least squares procedure.

In the general setting, for data given by tuples $(x_1, y_1), (x_2 y_2), \ldots (x_n, y_n)$, the researcher might expect the x and y values to be related by $y = b_1 f_1(x) + b_2 f_2(x) + \ldots + b_m f_m(x)$ where the constants b_i are to be determined and the functions $f_i(x)$ are the expected relationship between x and y values. The system of distances takes the form

$$\begin{cases} b_1 f_1(x_1) + b_2 f_2(x_1) + \ldots + b_m f_m(x_1) - y_1 \\ b_1 f_1(x_2) + b_2 f_2(x_2) + \ldots + b_m f_m(x_2) - y_2 \\ \vdots \\ b_1 f_1(x_n) + b_2 f_2(x_n) + \ldots + b_m f_m(x_n) - y_n \end{cases}$$

Set

$$A=\begin{bmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_m(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_m(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_n) & f_2(x_n) & \cdots & f_m(x_n) \end{bmatrix} \qquad B=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \qquad Y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

and again minimize $|AB-Y|$.

Least squares procedures are sophisticated mathematical curve fitting procedures used in all classical parametric methods (Thompson, 1994). Calculus and linear algebra tools are implemented throughout the procedure. For large amounts of data, computer algebra systems, such as MAPLE or statistics packages are useful for carrying out computations. Small data sets are helpful in making the linear least squares procedure concrete and enabling computations to be done by hand.

References

Hamilton, L. (1992). Regression with graphics: A second course in applied statistics. Pacific Grove: Brooks/Cole.

Seber, G.A.F. (1977). Linear regression analysis. New York: John Wiley & Sons, Inc.

Thompson, B. (1994, February). Why multivariate methods are usually vital is research: some basic concepts. Paper presented at the biennial meeting of the Southwestern Society for Research in Human Development, Austin, TX. (ERIC Document Reproduction Service No. ED 367 687)

Table 1
Data for example 1

Student	Number of absences (x_i)	Grade in the class	Ordered pair representation
1	2	B	(2,3)
2	2	B	(2,3)
3	5	C	(5,2)
4	2	C	(2,2)
5	3	C	(3,2)
6	0	A	(0,4)
7	3	C	(3,2)
8	8	D	(8,1)
9	0	A	(0,4)
10	1	A	(1,4)

Note: A = 4.0, B = 3.0, C = 2.0, D = 1.0, F = 0.0

Table 2
Computations for the formula to find b for example 1

$\sum_{i=1}^{n} x_i$	$\left(\sum_{i=1}^{n} x_i\right)^2$	$\sum_{i=1}^{n} y_i$	$\sum_{i=1}^{n} x_i y_i$	$\sum_{i=1}^{n} x_i^2$	$\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i$	$n\sum_{i=1}^{n} x_i y_i$	$n\sum_{i=1}^{n} x_i^2$
26	676	27	48	120	702	408	1200

Table 3
Data for example 2

Student	Math SAT score	H S Math Achievement Score	First Math Course Grade	College Grade Point Average	Ordered 4-tuples
1	670	10	A	3.82	(670,10,4,3.82)
2	630	9	A	3.74	(630,9,4,3.74)
3	610	6	D	1.84	(610,6,1,1.84)
4	570	9	B	3.34	(570,9,3,3.34)
5	700	8	B	3.26	(700,8,3,3.26)
6	640	6	C	2.35	(640,6,2,2.35)
7	630	7	C	3.03	(630,7,2,3.03)
8	610	8	B	3.05	(610,8,3,3.05)
9	570	10	A	3.76	(570,10,4,3.76)
10	550	8	C	2.72	(550,8,2,2.72)

Note: A = 4.0, B = 3.0, C = 2.0, D = 1.0, F = 0.0

Figure 1
Illustrates a line "closest" to four data points and the vertical distances whose sum of squares is to be minimized.

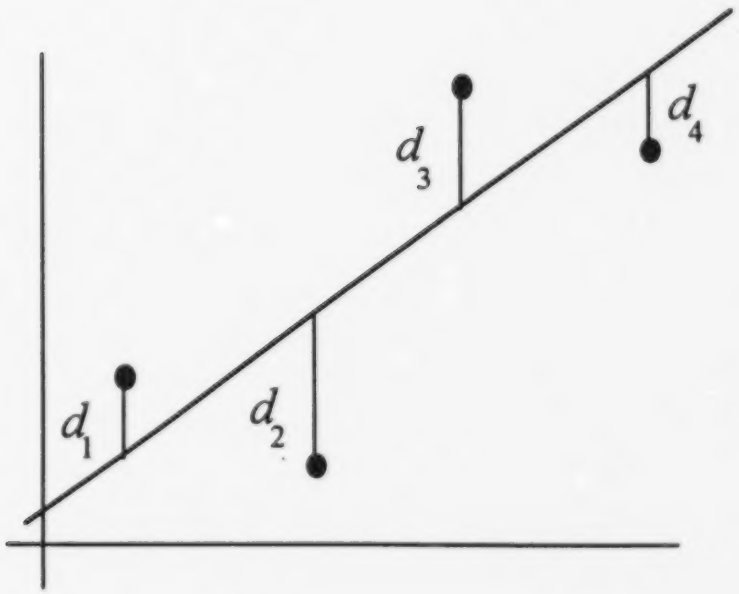
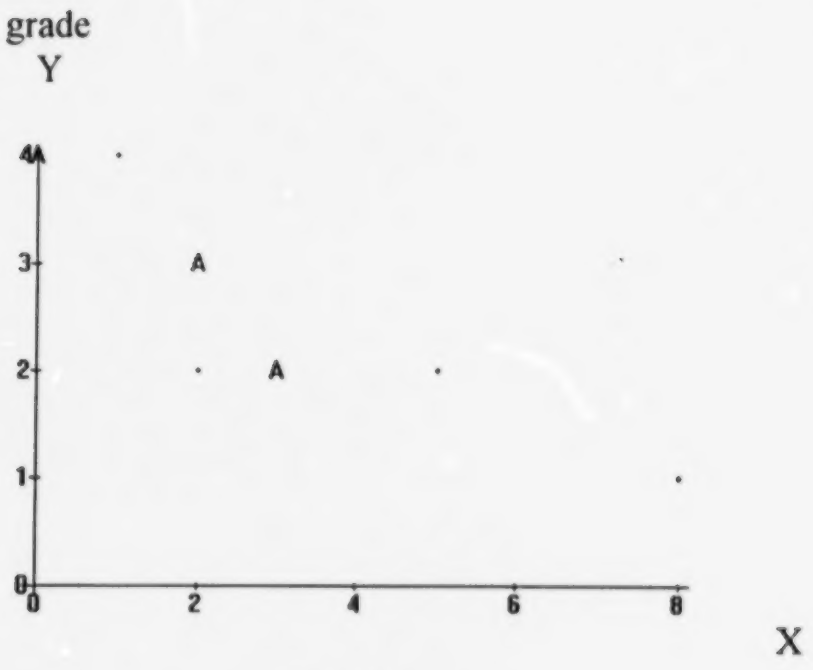
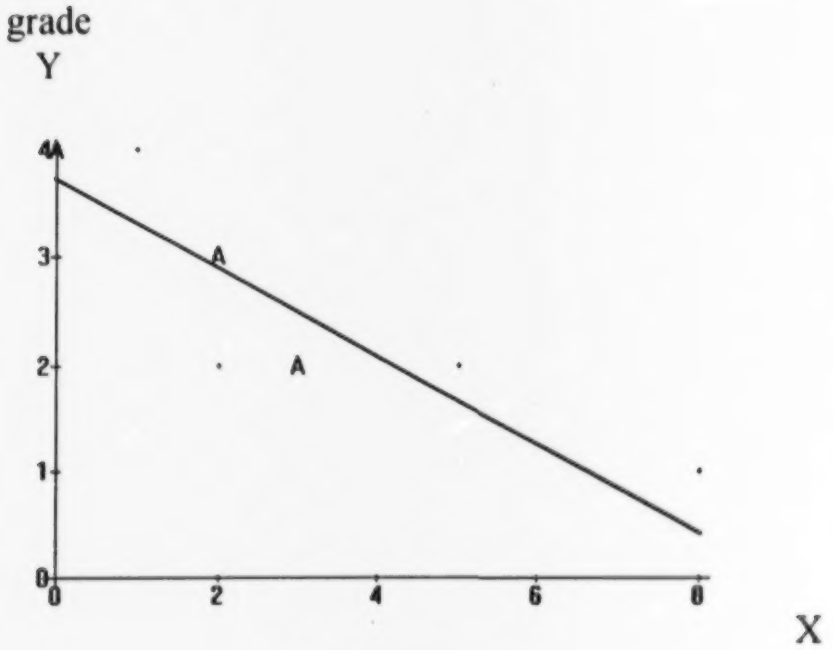


Figure 2
Scattergram of data for example 1



Note: "A" represents two points

Figure 3
Least Squares Line for data in example 1



Note: "A" denotes two points

APPENDIX

MAPLE Commands

```
>with(linalg):
>A:=matrix(#rows,#cols,[row entries separated by commas]);
>AT:=transpose(A);
>innerprod(A,AT);
>ATA:=innerprod(AT,A);
>ATAINV:=inverse(ATA);
>INVAT:=innerprod(ATAINV,AT);
>Y:=matrix(#rows,#cols,[row entries separated by commas]);
>B:=innerprod(INVAT,Y);
```